

PCT

INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/54, 9/12, 1/21, C07K 14/415, C12Q 1/68, 1/48	A2	(11) International Publication Number: WO 99/53075
		(43) International Publication Date: 21 October 1999 (21.10.99)

(21) International Application Number: PCT/US99/07638

(22) International Filing Date: 8 April 1999 (08.04.99)

(30) Priority Data:
60/081,132 9 April 1998 (09.04.98) US

(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KLEIN, Theodore, M. [US/US]; 2229 Rosewood Drive, Wilmington, DE 19810 (US). MORAKINYO, Layo, O. [US/US]; 216 Barrett Run Place, Newark, DE 19702 (US). ODELL, Joan, T. [US/US]; P.O. Box 826, Unionville, PA 19375 (US). SAKAI, Hajime [DE/US]; 105 Banbury Drive, Wilmington, DE 19803 (US).

(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: CELL CYCLE REGULATORY PROTEINS CDC-16, DP-1, DP-2 AND E2F FROM PLANTS

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a cell cycle regulatory protein. The invention also relates to the construction of a chimeric gene encoding all or a portion of the cell cycle regulatory protein, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the cell cycle regulatory protein in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY -

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

CELL CYCLE REGULATORY PROTEINS CDC-16, DP-1, DP-2 AND E2F FROM PLANTS

This application claims the benefit of U.S. Provisional Application No. 60/081,132, filed April 9, 1998.

5 FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding cell cycle regulatory proteins in plants and seeds.

10 BACKGROUND OF THE INVENTION

Cells divide by duplicating their chromosomes and segregating one copy of each duplicated chromosome, as well as providing essential organelles, to each of two daughter cells. Regulation of cell division is critical for the normal development of multicellular organisms. A cell that is destined to grow and divide must pass through specific phases of a cell cycle: G₁, S (period of DNA synthesis), G₂, and M (mitosis). Studies have shown that
15 cell division is controlled via the regulation of two critical events during the cell cycle: initiation of DNA synthesis and the initiation of mitosis. Several kinase proteins control cell cycle progression through these events. These protein kinases are heterodimeric proteins, having a cyclin-dependent kinase (Cdks) subunit and a cyclin subunit that provides the regulatory specificity to the heterodimeric protein. These heterodimeric proteins regulate
20 cell cycle by interacting with proteins involved in the initiation of DNA synthesis and mitosis and phosphorylating them at specific regulatory sites, activating some and inactivating others. The cyclin subunit concentration varies in phase with cell cycle while the concentration of the Cdks remain relatively constant throughout the cell cycle.

In mammalian cells transcription factor E2F genes encode a family of proteins that
25 bind to the nucleotide sequence TTTCGCGC and regulate the expression of various cellular and viral promoters. Another transcription factor family of proteins, DP-1 and DP-2, can form heterodimers with E2F proteins in vivo (Helin, K. et al. (1994) *Genes Dev.* 10:1850-1861; Ivey-Hoyl M. et al. (1993) *Mol. Cell Biol.* 13(2):7802-7812; and Zhang, Y. et al. (1995) *Oncogene* 10(11):2085-2093). The E2F-DP transcription factors are major
30 regulators of genes that are required for the progression of S-phase, such as DHFR and DNA polymerase alpha, and they play a critical role in cell cycle regulation and differentiation. The retinoblastoma tumor suppressor protein has been shown to induce growth arrest by binding to E2F-DP and repressing its activity. Lastly, CDC-16 is cell cycle protein identified in mammalian and yeast cells. This protein has been shown to localize to the
35 centrosome and mitotic spindle and is essential for the metaphase to anaphase transition during cell division (Lamb, J. R. et al. (1994) *EMBO J.* 13(18):4321-4328).

There is a great deal of interest in identifying the genes that encode cell cycle regulatory proteins in plants. These genes may be used to express cell cycle regulatory

SUBSTITUTE SHEET (RULE 26)

nucleic acid sequences encoding all or a portion of a cell cycle regulatory protein would facilitate studies to better understand cell cycle regulation in plants, provide genetic tools to enhance cell growth in tissue culture, increase the efficiency of gene transfer and help provide more stable transformations. Cell cycle regulatory proteins may also provide targets to facilitate design and/or identification of inhibitors of cell cycle regulatory proteins that may be useful as herbicides.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding cell cycle regulatory proteins. Specifically, this invention concerns an isolated nucleic acid fragment encoding a CDC-16, DP-1, DP-2 or E2F protein. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding a CDC-16, DP-1, DP-2 or E2F protein.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a cell cycle regulatory protein selected from the group consisting of CDC-16, DP-1, DP-2 and E2F.

In another embodiment, the instant invention relates to a chimeric gene encoding a CDC-16, DP-1, DP-2 or E2F protein, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding a CDC-16, DP-1, DP-2 or E2F protein, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a CDC-16, DP-1, DP-2 or E2F protein, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of a CDC-16, DP-1, DP-2 or E2F protein in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a CDC-16, DP-1, DP-2 or E2F protein; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of CDC-16, DP-1, DP-2 or E2F protein in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding a CDC-16, DP-1, DP-2 or E2F protein.

- 5 A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a CDC-16, DP-1, DP-2 or E2F protein, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a CDC-16, DP-1, DP-2 or E2F protein, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the
- 10 chimeric gene results in production of CDC-16, DP-1, DP-2 or E2F protein in the transformed host cell; (c) optionally purifying the CDC-16, DP-1, DP-2 or E2F protein expressed by the transformed host cell; (d) treating the CDC-16, DP-1, DP-2 or E2F protein with a compound to be tested; and (e) comparing the activity of the CDC-16, DP-1, DP-2 or E2F protein that has been treated with a test compound to the activity of an untreated
- 15 CDC-16, DP-1, DP-2 or E2F protein, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS

The invention can be more fully understood from the following detailed description and the accompanying Sequence Listing which form a part of this application.

- 20 The following sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence comprising a portion of the cDNA insert in clone cpf1c.pk001.k13 encoding a portion of a corn CDC-16 protein.

- 25 SEQ ID NO:2 is the deduced amino acid sequence of a portion of a CDC-16 protein derived from the nucleotide sequence of SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence comprising a portion of the cDNA insert in clone sfl1.pk0030.e3 encoding a portion of a soybean CDC-16 protein.

- 30 SEQ ID NO:4 is the deduced amino acid sequence of a portion of a CDC-16 protein derived from the nucleotide sequence of SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence comprising a portion of the cDNA insert in clone ids.pk0025.f7 encoding a portion of an *Impatiens balsamia* DP-1 protein.

SEQ ID NO:6 is the deduced amino acid sequence of a portion of a DP-1 protein derived from the nucleotide sequence of SEQ ID NO:5.

- 35 SEQ ID NO:7 is the nucleotide sequence comprising a portion of the cDNA insert in clone p0072.comfs64r encoding a portion of a corn DP-1 protein.

SEQ ID NO:8 is the deduced amino acid sequence of a portion of a DP-1 protein derived from the nucleotide sequence of SEQ ID NO:7.

SEQ ID NO:9 is the nucleotide sequence comprising a portion of the cDNA insert in clone src3c.pk018.m11 encoding a portion of a soybean DP-1 protein.

SEQ ID NO:10 is the deduced amino acid sequence of a portion of a DP-1 protein derived from the nucleotide sequence of SEQ ID NO:9.

5 SEQ ID NO:11 is the nucleotide sequence comprising a contig assembled from the cDNA inserts in clones p0005.cbmfh22r, cde1c.pk001.j13 and cen3n.pk0183.b9 encoding a portion of a corn DP-2 protein.

SEQ ID NO:12 is the deduced amino acid sequence of a portion of a DP-2 protein derived from the nucleotide sequence of SEQ ID NO:11.

10 SEQ ID NO:13 is the nucleotide sequence comprising the entire cDNA insert in clone wlmk1.pk0005.e2 encoding a portion of a wheat DP-2 protein.

SEQ ID NO:14 is the deduced amino acid sequence of a portion of a DP-2 protein derived from the nucleotide sequence of SEQ ID NO:13.

15 SEQ ID NO:15 is the nucleotide sequence comprising a portion of the cDNA insert in clone rsl1n.pk004.d15 encoding a portion of a rice E2F protein.

SEQ ID NO:16 is the deduced amino acid sequence of a portion of an E2F protein derived from the nucleotide sequence of SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising the entire cDNA insert in clone sel.pk0012.f4 encoding a portion of a soybean E2F protein.

20 SEQ ID NO:18 is the deduced amino acid sequence of a portion of an E2F protein derived from the nucleotide sequence of SEQ ID NO:17.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical*
25 *Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein,
30 an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide
35 sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence.

"Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the nucleic acid fragments disclosed herein.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent similarity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Preferred are those nucleic acid fragments whose nucleotide sequences encode amino acid sequences that are 80% similar to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are 90% similar to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% similar to the amino acid

sequences reported herein. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*.

- 5 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10) (hereafter Clustal algorithm). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

- 10 A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also
- 15 www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous
- 20 nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a
- 25 nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as
- 30 reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

- "Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that
- 35 encodes all or a substantial portion of the amino acid sequence encoding the CDC-16, DP-1, DP-2 or E2F proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16 and 18. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for

improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

“Synthetic genes” can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. “Chemically synthesized”, as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. “Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a DNA sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a

promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology* 3:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed.

"Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation

technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several cell cycle regulatory proteins have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1
Cell Cycle Regulatory Proteins

Enzyme	Clone	Plant
CDC-16	cpf1c.pk001.k13	Corn
	sfl1.pk0030.e3	Soybean
DP-1	ids.pk0025.f7	<i>Impatiens</i>
	p0072.comfs64r	Corn
	src3c.pk018.m11	Soybean
DP-2	p0005.cbmfh22r	Corn
	cde1c.pk001.j13	Corn
	cen3n.pk0183.b9	Corn
	wlmk1.pk0005.e2	Wheat
E2F	rsl1n.pk004.d15	Rice
	sel.pk0012.f4	Soybean

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other CDC-16, DP-1, DP-2 or E2F proteins, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific

oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *PNAS USA* 85:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *PNAS USA* 86:5673; Loh et al., (1989) *Science* 243:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M. A. and Martin, G. R., (1989) *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R. A. (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed CDC-16, DP-1, DP-2 or E2F proteins are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering cell cycle regulation those cells.

Overexpression of the CDC-16, DP-1, DP-2 or E2F proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues

at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J.* 4:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

It may also be desirable to reduce or eliminate expression of genes encoding CDC-16, DP-1, DP-2 or E2F proteins in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant cell cycle regulatory proteins can be constructed by linking a gene or gene fragment encoding a CDC-16, DP-1, DP-2 or E2F protein to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant CDC-16, DP-1, DP-2 or E2F proteins (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting CDC-16, DP-1, DP-2 or E2F proteins *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant CDC-16, DP-1, DP-2 or E2F proteins are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant CDC-16, DP-1, DP-2 or E2F proteins. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded cell cycle regulatory protein. An example of a vector for high level expression of the instant CDC-16, DP-1, DP-2 or E2F proteins in a bacterial host is provided (Example 9).

Additionally, the instant CDC-16, DP-1, DP-2 or E2F proteins can be used as a targets to facilitate design and/or identification of inhibitors of those proteins that may be useful as herbicides. This is desirable because the CDC-16, DP-1, DP-2 or E2F proteins described herein are involved in the regulation of cell cycle. Accordingly, inhibition of the activity of one or more of the proteins described herein could lead to inhibition plant growth. Thus, the instant CDC-16, DP-1, DP-2 or E2F proteins could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al., (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter* 4(1):37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med.* 114(2):95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA* 92:8149; Bensen et al., (1995) *Plant Cell* 7:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the CDC-16, DP-1, DP-2 or E2F protein. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a CDC-16, DP-1, DP-2 or E2F protein can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the CDC-16, DP-1, DP-2 or E2F protein gene product.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without

departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

- 5 cDNA libraries representing mRNAs from various corn, *Impatiens*, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, *Impatiens*, Rice, Soybean and Wheat

Library	Tissue	Clone
cde1c	Corn developing embryo 20 days after pollination	cde1c.pk001.j13
cen3n	Corn endosperm days after pollination*	cen3n.pk0183.b9
cpflc	Corn tissue treated with chemicals related to protein synthesis and pooled***	cpflc.pk001.k13
ids	<i>Impatiens balsamina</i> developing seed	ids.pk0025.f7
p0005	Corn immature ear	p0005.cbmfh22r
p0072	Corn 14 days after planting etiolated seedling: Mesocotyl	p0072.comfs64r
rsl1n	Rice 15 day old seedling*	rsl1n.pk004.d15
se1	Soybean embryo, 6-10 day after flowering	se1.pk0012.f4
sfl1	Soybean immature flower	sfl1.pk0030.e3
src3c	Soybean 8 day old root inoculated with eggs of cyst nematode <i>Heterodera glycines</i> (Race 14) for 4 days.	src3c.pk018.m11
wlmk1	Wheat seedlings 1 hr after inoculation with <i>Erysiphe graminis f. sp tritici</i> and treatment with fungicide**	wlmk1.pk0005.e2

- 10 *These libraries were normalized essentially as described in U.S. Patent No. 5,482,845

**Application of 6-iodo-2-propoxy-3-propyl-4(3H)-quinazolinone; synthesis and methods of using this compound are described in USSN 08/545,827, incorporated herein by reference.

- 15 ***Chloramphenicol, cyclohexamide, neomycin sulfate and aurintricarboxylic acid commercially available from Sigma Chemical Co. and Calbiochem-Novabiochem Corp Calbiochem.

- cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences

(expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science* 252:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding cell cycle regulatory proteins were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272 and Altschul, Stephen F., et al. (1997) *Nucleic Acids Res.* 25:3389-3402) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding CDC-16 Proteins

The BLASTX search using the EST sequences from clones cpflc.pk001.k13 and sfl1.pk0030.e3 revealed similarity of the proteins encoded by the cDNAs to CDC-16 protein from *Homo sapiens* (NCBI Identifier No. gi 1362769). The BLAST results for each of these ESTs are shown in Table 3:

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous
to *Homo sapiens* CDC-16 Protein

Clone	BLAST pLog Score
cpflc.pk001.k13	11.00
sfl1.pk0030.e3	50.40

The sequence of a portion of the cDNA insert from clone cpflc.pk001.k13 is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA, which represents 14% of the N-terminal region of the protein, is shown in SEQ ID NO:2. A calculation of the percent

similarity of the amino acid sequence set forth in SEQ ID NO:2 and the *Homo sapiens* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:2 is 40% similar to the *Homo sapiens* CDC-16 protein.

The sequence of a portion of the cDNA insert from clone sfl1.pk0030.e3 is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA, which represents 43% of the protein (middle region), is shown in SEQ ID NO:4. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:4 and the *Homo sapiens* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:4 is 46% similar to the *Homo sapiens* CDC-16 protein. The percent similarity between the corn and soybean amino acid sequence was calculated to be 18% using the Clustal algorithm.

BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of CDC-16 proteins. These sequences represent the first plant sequences encoding CDC-16 proteins.

EXAMPLE 4

15 Characterization of cDNA Clones Encoding DP-1 Protein

The BLASTX search using the EST sequence from clone ids.pk0025.f7 revealed similarity of the protein encoded by the cDNA to DP-1 protein from *Xenopus laevis* (NCBI Identifier No. gi 913227). The BLASTX search using the EST sequences from clones p0072.comfs64r and src3c.pk018.m11 revealed similarity of the proteins encoded by the cDNAs to DP-1 protein from *Mus musculus* (NCBI Identifier No. gi 420232).

The BLAST results for each of these ESTs are shown in Table 4:

TABLE 4

25 BLAST Results for Clones Encoding Polypeptides Homologous to *Xenopus laevis* and *Mus musculus* DP-1 Protein

Clone	BLAST pLog Score
ids.pk0025.f7	23.30
p0072.comfs64r	5.00
src3c.pk018.m11	17.00

The sequence of a portion of the cDNA insert from clone ids.pk0025.f7 is shown in SEQ ID NO:5; the deduced amino acid sequence of this cDNA, which represents 33% of the of the protein (middle region), is shown in SEQ ID NO:6. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:8 and the *Xenopus laevis* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:6 is 46% similar to the *Xenopus laevis* DP-1 protein.

The sequence of a portion of the cDNA insert from clone p0072.comfs64r is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA, which represents 23% of the protein (middle region), is shown in SEQ ID NO:8. A calculation of the percent similarity

of the amino acid sequence set forth in SEQ ID NO:8 and the *Mus musculus* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:8 is 37% similar to the *Mus musculus* DP-1 protein.

5 The sequence of a portion of the cDNA insert from clone src3c.pk018.m11 is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA, which represents 42% of the protein (middle region), is shown in SEQ ID NO:10. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:10 and the *Mus musculus* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:10 is 48% similar to the *Mus musculus* DP-1 protein.

10 The percent similarity between the corn, *Impatiens* and soybean amino acid sequence was calculated to range between 31 to 78% using the Clustal algorithm.

BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of DP-1 proteins. These sequences represent the first plant sequences encoding DP-1 proteins.

15

EXAMPLE 5

Characterization of cDNA Clones Encoding DP-2 Protein

The BLASTX search using the EST sequences from clones p0005.cbmfh22r, cde1c.pk001.j13 and cen3n.pk0183.b9 revealed similarity of the proteins encoded by the cDNAs to DP-2 protein from *Homo sapiens* (NCBI Identifier No. gi 604479). The
20 BLASTX search using the EST sequence from clone wlmk1.pk0005.e2 revealed similarity of the protein encoded by the cDNA to DP-2 protein from *Mus* species (NCBI Identifier No. gi 3122929).

In the process of comparing the ESTs it was found that corn clones p0005.cbmfh22r, cde1c.pk001.j13 and cen3n.pk0183.b9 had overlapping regions of homology. Using this
25 homology it was possible to align the ESTs and assemble a contig encoding a unique corn DP-2 protein.

The BLAST results for the corn contig and the wheat EST are shown in Table 5:

TABLE 5

30 BLAST Results for Clones Encoding Polypeptides Homologous to *Homo sapiens* and *Mus* species DP-2 Protein

Clone	BLAST pLog Score
Contig composed of: p0005.cbmfh22r cde1c.pk001.j13 cen3n.pk0183.b9	45.10
wlmk1.pk0005.e2	7.00

The sequence of the corn contig composed of clones p0005.cbmfh22r, cde1c.pk001.j13 and cen3n.pk0183.b9 is shown in SEQ ID NO:11; the deduced amino acid

sequence of this contig, which represents 50% of the protein (middle region), is shown in SEQ ID NO:12. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:12 and the *Homo sapiens* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:8 is 52% similar to the *Homo sapiens* DP-2 protein.

5 The sequence of the entire cDNA insert from clone wlmk1.pk0005.e2 is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA, which represents 25% of the of the protein (middle region), is shown in SEQ ID NO:14. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:14 and the *Mus* species sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:14
10 is 44% similar to the *Mus* species DP-2 protein.

The percent similarity between the corn and wheat amino acid sequence was calculated to be 86% using the Clustal algorithm.

BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of DP-2 proteins. These sequences represent the first plant sequences encoding
15 DP-2 proteins.

EXAMPLE 6

Characterization of cDNA Clones Encoding E2F Proteins

The BLASTX search using the EST sequence from clone rsl1n.pk004.d15 revealed similarity of the protein encoded by the cDNA to E2F-4 protein from *Homo sapiens* (NCBI
20 Identifier No. gi 1061146). The BLASTX search using the EST sequence from clone sel.pk0012.f4 revealed similarity of the protein encoded by the cDNA to E2F protein from *Drosophila melanogaster* (NCBI Identifier No. gi 3551069).

The BLAST results for each of the ESTs are shown in Table 6:

25

TABLE 6

BLAST Results for Clones Encoding Polypeptides Homologous
to *Homo sapiens* and *Drosophila melanogaster* E2F Proteins

Clone	BLAST pLog Score
rsl1n.pk004.d15	5.52
sel.pk0012.f4	12.00

30 The sequence of a portion of the cDNA insert from clone rsl1n.pk004.d15 is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA, which represents 12% of the of the protein (middle region), is shown in SEQ ID NO:16. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:16 and the *Homo sapiens* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:16 is 49% similar to the *Homo sapiens* E2F-4 protein.

35 The sequence of the entire cDNA insert from clone sel.pk0012.f4 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA, which represents 10% of the of the

protein (middle region), is shown in SEQ ID NO:18. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:18 and the *Drosophila melanogaster* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:18 is 43% similar to the *Drosophila melanogaster* E2F protein.

- 5 The percent similarity between the rice and soybean amino acid sequence was calculated to be 15% using the Clustal algorithm.

BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of E2F proteins. These sequences represent the first plant sequences encoding E2F proteins.

10

EXAMPLE 7

Expression of Chimeric Genes in Monocot Cells

- A chimeric gene comprising a cDNA encoding a cell cycle regulatory protein in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be
15 constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested
20 with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103
25 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by
30 restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a cell cycle regulatory protein, and the 10 kD zein 3' region.

- 35 The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu

et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-

supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

- Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

EXAMPLE 8

Expression of Chimeric Genes in Dicot Cells

- A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant cell cycle regulatory protein in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

- The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

- Soybean embryos may then be transformed with the expression vector comprising a sequence encoding a cell cycle regulatory protein. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

- Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Kline et al. (1987) *Nature* (London) 327:70, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the cell cycle regulatory protein and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 μ L of a 60 mg/mL 1 μ m gold particle suspension is added (in order): 5 μ L DNA (1 μ g/ μ L), 20 μ L spermidine (0.1 M), and 50 μ L CaCl_2 (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 μ L 70% ethanol and resuspended in 40 μ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five μ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane-rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 9

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant cell cycle regulatory proteins can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM

with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

- 5 Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the
- 10 manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized
- 15 with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the cell cycle regulatory protein are then screened for the correct orientation with respect to the T7 promoter by
- 20 restriction enzyme analysis.

- For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately
- 25 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe
- 30 sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 10

- 35 Evaluating Compounds for Their Ability to Inhibit the Activity
of Cell Cycle Regulatory Proteins

The cell cycle regulatory proteins described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 9, or expression in eukaryotic cell culture,

in planta, and using viral expression systems in suitably infected organisms or cell lines.

The instant cell cycle regulatory proteins may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione

- 5 S-transferase ("GST"); thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("His₆"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically
- 10 cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant cell cycle regulatory proteins, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH

- 15 precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the cell cycle regulatory proteins are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, a cell cycle regulatory protein may be expressed as a
- 20 fusion protein coupled to the C-terminus of thioredoxin. In addition, a (His)₆ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate
- 25 embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β -mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may
- 30 be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified protein, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the cell cycle regulatory proteins disclosed herein. Assays may be conducted

35 under well known experimental conditions which permit optimal protein activity. For example, assays for E2F and DP-1 transcription factor activities are presented by Helin K., et al., 1994 *Genes Dev.* 10:1850-1861 and Ivey-Hoyl M., et al., 1993 *Mol. Cell Biol.* 13(2):7802-7812. Assays for DP-2 activity are presented by Zhang Y., et al., 1995

Oncogene 10(11):2085-2093. Assays for CDC-16 activity are presented by Lamb, J. R., et al., 1994 *EMBO J.* 13(18):4321--4328.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of a CDC-16 protein comprising a member selected from the group consisting of:
 - 5 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2 and 4;
 - (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2 and 4; and
 - 10 (c) an isolated nucleic acid fragment that is complementary to (a) or (b).
2. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1 and 3.
- 15 3. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.
4. A transformed host cell comprising the chimeric gene of Claim 3.
5. A CDC-16 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2 and 4.
- 20 6. An isolated nucleic acid fragment encoding all or a substantial portion of a DP-1 protein comprising a member selected from the group consisting of:
 - 25 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:6, 8 and 10;
 - (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:6, 8 and 10; and
 - 30 (c) an isolated nucleic acid fragment that is complementary to (a) or (b).
7. The isolated nucleic acid fragment of Claim 6 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:5, 7 and 9.
- 35 8. A chimeric gene comprising the nucleic acid fragment of Claim 6 operably linked to suitable regulatory sequences.
9. A transformed host cell comprising the chimeric gene of Claim 8.

10. A DP-1 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:6, 8 and 10.

5 11. An isolated nucleic acid fragment encoding all or a substantial portion of a DP-2 protein comprising a member selected from the group consisting of:

- (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:12 and 14;
- 10 (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:12 and 14; and
- (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

12. The isolated nucleic acid fragment of Claim 11 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:11 and 13.

13. A chimeric gene comprising the nucleic acid fragment of Claim 11 operably linked to suitable regulatory sequences.

14. A transformed host cell comprising the chimeric gene of Claim 13.

20 15. A DP-2 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:12 and 14.

16. An isolated nucleic acid fragment encoding all or a substantial portion of a E2F protein comprising a member selected from the group consisting of:

- 25 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:16 and 18;
- (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:16 and 18; and
- 30 (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

17. The isolated nucleic acid fragment of Claim 16 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:15 and 17.

35 18. A chimeric gene comprising the nucleic acid fragment of Claim 16 operably linked to suitable regulatory sequences.

19. A transformed host cell comprising the chimeric gene of Claim 18.

20. An E2F polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:16 and 18.

21. A method of altering the level of expression of a cell cycle regulatory protein in a host cell comprising:

- (a) transforming a host cell with the chimeric gene of any of Claims 3, 8, 13 and 18; and
- 5 (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene

wherein expression of the chimeric gene results in production of altered levels of a cell cycle regulatory protein in the transformed host cell.

22. A method of obtaining a nucleic acid fragment encoding all or a substantial
10 portion of the amino acid sequence encoding a cell cycle regulatory protein comprising:

- (a) probing a cDNA or genomic library with the nucleic acid fragment of any of Claims 1, 6, 11 and 16 ;
- (b) identifying a DNA clone that hybridizes with the nucleic acid fragment of any of Claims 1, 6, 11 and 16;
- 15 (c) isolating the DNA clone identified in step (b); and
- (d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a cell cycle regulatory protein.

23. A method of obtaining a nucleic acid fragment encoding a substantial portion
20 of an amino acid sequence encoding a cell cycle regulatory protein comprising:

- (a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, and 17; and
- 25 (b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a cell cycle regulatory protein.

30 24. The product of the method of Claim 22.

25. The product of the method of Claim 23.

26. A method for evaluating at least one compound for its ability to inhibit the activity of a cell cycle regulatory protein, the method comprising the steps of:

- 35 (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a cell cycle regulatory protein operably linked to suitable regulatory sequences;
- (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the cell cycle regulatory protein encoded

- by the operably linked nucleic acid fragment in the transformed host cell;
- 5 (c) optionally purifying the cell cycle regulatory protein expressed by the transformed host cell;
- (d) treating the cell cycle regulatory protein with a compound to be tested; and
- (e) comparing the activity of the cell cycle regulatory protein that has been treated with a test compound to the activity of an untreated cell cycle regulatory protein,
- 10 thereby selecting compounds with potential for inhibitory activity.

SEQUENCE LISTING

<110> E. I. DU PONT DE NEMOURS AND COMPANY

<120> CELL CYCLE REGULATORY PROTEINS

<130> BB-1161

<140>

<141>

<150> 60/081,132

<151> APRIL 9, 1998

<160> 18

<170> Microsoft Office 97

<210> 1

<211> 507

<212> DNA

<213> Zea mays

<220>

<221> unsure

<222> (333)

<220>

<221> unsure

<222> (364)

<220>

<221> unsure

<222> (403)

<220>

<221> unsure

<222> (426)

<220>

<221> unsure

<222> (482)

<400> 1

```

gagaccaccg atccgccgct ggagatgagg gaggaggcgc tggagcggct gcgcgggggtg 60
gtacggggaca gcgccgggaa gcacctctac acgtcggcca tcttcctcgc cgacaagggtg 120
gcggcgggcca cgggggaccc tggcgacgct tacatgctcg cgcaggcact cttcctcggg 180
cgtcaattcc gtcacgcgct ccacctctc aacaattctc gcctgctcgc cgacctccgg 240
ttcagattcc tcgccgcca gtgcctcgag gagttgaaag agtggcatca gtgtttgttg 300
atgcttgggg atgcaaaagt ggatgagcat ggnaatgtcc tttgatcatg atgatgacag 360
tganatttat ttgataagg acgcggaaag atcatgagat cantatcaaa tcagctctat 420
gtttcntacg tggtaaaggc atatgaagca ctgggacaac cgtgatcttg ctcgccaatg 480
gnacaaaggc agctgttaaa gccgatg 507

```

<210> 2

<211> 89

<212> PRT

<213> Zea mays

<400> 2

```

Leu Glu Arg Leu Arg Gly Val Val Arg Asp Ser Ala Gly Lys His Leu
  1             5             10             15

```

Tyr Thr Ser Ala Ile Phe Leu Ala Asp Lys Val Ala Ala Ala Thr Gly
 20 25 30
 Asp Pro Gly Asp Val Tyr Met Leu Ala Gln Ala Leu Phe Leu Gly Arg
 35 40 45
 Gln Phe Arg His Ala Leu His Leu Leu Asn Asn Ser Arg Leu Leu Arg
 50 55 60
 Asp Leu Arg Phe Arg Phe Leu Ala Ala Lys Cys Leu Glu Glu Leu Lys
 65 70 75 80
 Glu Trp His Gln Cys Leu Leu Met Leu
 85

<210> 3
 <211> 1093
 <212> DNA
 <213> Glycine max

<400> 3
 gcacgaggtt ttgagttaac aaatgatttg cttgagaagg atctttttcca tctaaagact 60
 acgctagtag atctggcagc tgctgtggag cttgggcatt caaatgaact ctatctgatg 120
 tcatgcaatt tgggtgaagga ttatcctcag atggccttat catggtttgc tggttggttgc 180
 tactactatt gtatcaagaa gtatgatcaa tcccgccgtt atttttagcaa ggcaactagt 240
 ctagatggaa catttcccc tgcatggata ggctatggga atgcttatgc tgctcaagaa 300
 gagggagacc aggctatgtc tgcttaccgc actgcagcta gattgtttcc tgggtgccat 360
 ttagcaactc ttacattgg aatggaatgc atgcggaccc acagttataa gcttgctgag 420
 cagtttttta cgcaggccaa gtcaatttgc tcttcagatc cacttgtgta taatgaactt 480
 ggagttgttg cctaccatat ggaagagtat aagaaagcag tgtggtggtt tgagaaaaca 540
 ttagccttgg taccaaccac ttatctgaa atttggaat cgactgtagt caatcttgca 600
 catgcataca gaaagctgaa gatgtaccga gaagcaattt catattatga gaaagcactt 660
 gcattatcaa caagaagtgt gagcacttat gccggtcttg catatactta ccacctacag 720
 gatgatttca ccacagcaat tgcattattat cacaagcct tgtggctgaa accagatgat 780
 cagttctgca ctgaaatgtt aagttgggct ctaatatagtg aaagtcgaag aggcgttgac 840
 cccagtcctt aattccgttg aagtcctgatt tgaatatgat ggtgcagcca ttgacattg 900
 taaatgtgta acaatgagtg ccgagttgat gtattgccaa tgtgcaattg ttaattatta 960
 ttatcagttg tagataatct gtattttgtt tctgaataat gtttgatttt attatatctc 1020
 tagcaatttt atactagtta aatattccgt gacttcacag gtattaatct taaaatttgt 1080
 aatttttaatt att 1093

<210> 4
 <211> 264
 <212> PRT
 <213> Glycine max

<400> 4
 Phe Glu Leu Thr Asn Asp Leu Leu Glu Lys Asp Leu Phe His Leu Lys
 1 5 10 15
 Thr Thr Leu Val His Leu Ala Ala Ala Val Glu Leu Gly His Ser Asn
 20 25 30
 Glu Leu Tyr Leu Met Ser Cys Asn Leu Val Lys Asp Tyr Pro Gln Met
 35 40 45
 Ala Leu Ser Trp Phe Ala Val Gly Cys Tyr Tyr Tyr Cys Ile Lys Lys
 50 55 60
 Tyr Asp Gln Ser Arg Arg Tyr Phe Ser Lys Ala Thr Ser Leu Asp Gly
 65 70 75 80
 Thr Phe Pro Pro Ala Trp Ile Gly Tyr Gly Asn Ala Tyr Ala Ala Gln
 85 90 95

Glu Glu Gly Asp Gln Ala Met Ser Ala Tyr Arg Thr Ala Ala Arg Leu
 100 105 110
 Phe Pro Gly Cys His Leu Ala Thr Leu Tyr Ile Gly Met Glu Cys Met
 115 120 125
 Arg Thr His Ser Tyr Lys Leu Ala Glu Gln Phe Phe Thr Gln Ala Lys
 130 135 140
 Ser Ile Cys Ser Ser Asp Pro Leu Val Tyr Asn Glu Leu Gly Val Val
 145 150 155 160
 Ala Tyr His Met Glu Glu Tyr Lys Lys Ala Val Trp Trp Phe Glu Lys
 165 170 175
 Thr Leu Ala Leu Val Pro Thr Thr Leu Ser Glu Ile Trp Glu Ser Thr
 180 185 190
 Val Val Asn Leu Ala His Ala Tyr Arg Lys Leu Lys Met Tyr Arg Glu
 195 200 205
 Ala Ile Ser Tyr Tyr Glu Lys Ala Leu Ala Leu Ser Thr Arg Ser Val
 210 215 220
 Ser Thr Tyr Ala Gly Leu Ala Tyr Thr Tyr His Leu Gln Asp Asp Phe
 225 230 235 240
 Thr Thr Ala Ile Ala Tyr Tyr His Lys Ala Leu Trp Leu Lys Pro Asp
 245 250 255
 Asp Gln Phe Cys Thr Glu Met Leu
 260

<210> 5
 <211> 665
 <212> DNA
 <213> Impatiens balsamia

<220>
 <221> unsure
 <222> (199)

<220>
 <221> unsure
 <222> (414)

<220>
 <221> unsure
 <222> (416)

<220>
 <221> unsure
 <222> (478)

<220>
 <221> unsure
 <222> (519)

<220>
 <221> unsure
 <222> (548)

<220>
 <221> unsure
 <222> (554)

<220>
 <221> unsure
 <222> (569)

<220>
 <221> unsure
 <222> (578)

<220>
 <221> unsure
 <222> (598)

<220>
 <221> unsure
 <222> (611)

<220>
 <221> unsure
 <222> (617)..(618)

<220>
 <221> unsure
 <222> (632)

<220>
 <221> unsure
 <222> (635)

<220>
 <221> unsure
 <222> (637)

<220>
 <221> unsure
 <222> (647)

<220>
 <221> unsure
 <222> (657)

<400> 5

agaagaaggg	tatatgatgc	actcaatgta	ctcatggcca	tggatataat	atccaaggac	60
aaaaaggaaa	ttcaatggaa	gggtcttcca	aggactagct	tgaatgacat	tgaagagata	120
aaggcagagc	gtctagggct	gagaagcaga	attgacaaaa	aaacagctta	tttgcaagaa	180
cttgaggaac	attatgtang	tcttcagaat	cttgtacaaa	ggaatgagcg	gttatatggc	240
tcaggggaaca	tgccgactgg	aggtgtggct	ttgcctttta	tacttgtgca	gacacgaccc	300
catgcaaccg	tcgaaataga	aatatcggaa	gatatgcagc	tggttcattt	cgacttcaac	360
agcactcctt	ttgagctaca	cgacgataat	tatgttatga	gggcaatgaa	attnancgga	420
ataaacgatg	gtgaatcaat	gccaacttac	cgacgatgga	ggtgaaggcc	aagtattnca	480
agccaacgtc	aaggcaaacg	aaattgcgtg	ctctcttcnt	aacttcgaca	tagaatttgt	540
tcgtttgnca	cacnctaaat	gctggtcgnt	ggtctgtnac	tgtaacgaaa	tccgttantt	600
gatattctgg	naaaatnntg	caggattggg	gnganancca	cctgaancta	accttancag	660
gattg						665

<210> 6
 <211> 138
 <212> PRT
 <213> Impatiens balsamia

<220>
 <221> UNSURE
 <222> (67)

<220>
 <221> UNSURE
 <222> (86)..(87)

<400> 6
 Arg Arg Arg Val Tyr Asp Ala Leu Asn Val Leu Met Ala Met Asp Ile
 1 5 10 15
 Ile Ser Lys Asp Lys Lys Glu Ile Gln Trp Lys Gly Leu Pro Arg Thr
 20 25 30
 Ser Leu Asn Asp Ile Glu Glu Ile Lys Ala Glu Arg Leu Gly Leu Arg
 35 40 45
 Ser Arg Ile Asp Lys Lys Thr Ala Tyr Leu Gln Glu Leu Glu Glu His
 50 55 60
 Tyr Val Xaa Leu Gln Asn Leu Val Gln Arg Asn Glu Arg Leu Tyr Gly
 65 70 75 80
 Ser Gly Asn Met Pro Xaa Xaa Pro Thr Gly Gly Val Ala Leu Pro Phe
 85 90 95
 Ile Leu Val Gln Thr Arg Pro His Ala Thr Val Glu Ile Glu Ile Ser
 100 105 110
 Glu Asp Met Gln Leu Val His Phe Asp Phe Asn Ser Thr Phe Glu Leu
 115 120 125
 His Asp Asp Asn Tyr Val Met Arg Ala Met
 130 135

<210> 7
 <211> 296
 <212> DNA
 <213> Zea mays

<400> 7
 gatattgaag aattgaagac tgagcttggt ggactgaaag gtagaattga gaagaaaagt 60
 gtttacttac aggagctaca agatcaatat gtaggtttgc aaaacctgat tcaacgaaat 120
 gagcaattat atggttcagg aaacacacct tctgggtggag tggctttgcc attcatccta 180
 gtccagaccc gacctcatgc aaccgtggaa gttgagatat cagaagatat gcagctggtt 240
 cattttgact tcaatagcac tccatttgag ctgcatgatg actcatatgt cctaaa 296

<210> 8
 <211> 100
 <212> PRT
 <213> Zea mays

<220>
 <221> UNSURE
 <222> (51)..(52)

<400> 8
 Asp Ile Glu Glu Leu Lys Thr Glu Leu Val Gly Leu Lys Gly Arg Ile
 1 5 10 15
 Glu Lys Lys Ser Val Tyr Leu Gln Glu Leu Gln Asp Gln Tyr Val Gly
 20 25 30

<220>
<221> unsure
<222> (195)

<220>
<221> unsure
<222> (214)

<220>
<221> unsure
<222> (219)

<220>
<221> unsure
<222> (224)

<220>
<221> unsure
<222> (247)

<220>
<221> unsure
<222> (265)

<220>
<221> unsure
<222> (283)

<220>
<221> unsure
<222> (289)

<220>
<221> unsure
<222> (301)

<220>
<221> unsure
<222> (315) .. (316)

<220>
<221> unsure
<222> (323)

<220>
<221> unsure
<222> (359)

<220>
<221> unsure
<222> (361)

<220>
<221> unsure
<222> (376)

<220>
<221> unsure
<222> (407)

<220>
<221> unsure
<222> (426) .. (427) .. (428)

<220>
 <221> unsure
 <222> (445)

<220>
 <221> unsure
 <222> (448)

<220>
 <221> unsure
 <222> (457)

<220>
 <221> unsure
 <222> (460)

<400> 9
 agcaacaata tgatgagaaa aacatccgcc gaagggncga tgatgctctg aacgtttctca 60
 tggcaatgga tattatttct antgncaaaa aaaaaattca atggaggggn cntcctcnca 120
 ctactgtgaa tgatattgaa gaactaaaga canancgtct tgggctcang natagaattg 180
 aanagaaaaa aaccnatctg cacgagcttg agngcaant cgtntgtctt cacaacctta 240
 ttcaacnaaa tgagcaagtt atatngctca agaaatcctc ccncccgng ggggggcctt 300
 nccttttttt tggtnnagaa aantcccat gcaactgtgg ggggggggaa tatcaaaaana 360
 natgcaacct ggtcantttg atttcaataa aaaacccttt ttttgcntgg cgaaaattaa 420
 tttttnnncg caaggggaatt ttgtntgnga ccactgnccn gggtaatatg acacaaaaac 480
 c 481

<210> 10
 <211> 83
 <212> PRT
 <213> Glycine max

<220>
 <221> UNSURE
 <222> (10)

<220>
 <221> UNSURE
 <222> (25) .. (26)

<220>
 <221> UNSURE
 <222> (35)

<220>
 <221> UNSURE
 <222> (37)

<220>
 <221> UNSURE
 <222> (49)

<220>
 <221> UNSURE
 <222> (54) .. (55)

<220>
 <221> UNSURE
 <222> (59)

<220>
 <221> UNSURE
 <222> (63)

<220>
 <221> UNSURE
 <222> (69)

<220>
 <221> UNSURE
 <222> (71)

<220>
 <221> UNSURE
 <222> (80)

<400> 10
 Tyr Asp Glu Lys Asn Ile Arg Arg Arg Xaa Asp Asp Ala Leu Asn Val
 1 5 10 15
 Leu Met Ala Met Asp Ile Ile Ser Xaa Xaa Lys Lys Lys Ile Gln Trp
 20 25 30
 Arg Gly Xaa Pro Xaa Thr Thr Val Asn Asp Ile Glu Glu Leu Lys Thr
 35 40 45
 Xaa Arg Leu Gly Leu Xaa Xaa Arg Ile Glu Xaa Lys Thr Thr Xaa Leu
 50 55 60
 His Glu Leu Glu Xaa Gln Xaa Val Cys Leu His Asn Leu Ile Gln Xaa
 65 70 75 80
 Asn Glu Gln

<210> 11
 <211> 1193
 <212> DNA
 <213> Zea mays

<220>
 <221> unsure
 <222> (238)

<400> 11
 ggcagacgac gttcctccgc ttgaataatc tcgacatcaa cggcgacgac ggcgcgtcgt 60
 cgcaggctcc tacgagcaag aagaaaagga gaggcacacg ggcagtgagg cctgataaag 120
 gtaaccgggg actgcgccag tttagtagtg aagtttgtga gaaagttgaa agtaaaggga 180
 gaacaacata taatgaggtg gcagatgaac ttgttgctga gtttacagac cccaacanta 240
 atattgaggg accagatcct gataacccta acgcgeaaca atatgatgag aaaaatatac 300
 gagggcgagt ttatgatgct ttaaattgtt tgatggctat ggacattata tctaaagata 360
 aaaaggagat ccagtggag ggcttgccgc ggactagtag aagtgatatt gaagaattga 420
 agactgagct tgtgggactg aaaggtagaa ttgagaagaa aagtgtttac ttacaggagc 480
 tacaagatca atatgtaggt ttgcaaaacc tgattcaacg aaatgagcaa ttatatggtt 540
 caggaaacac accttctggt ggagtggctt tgccattcat cctagtccag acccgacctc 600
 atgctaccgt ggaagttag atatacagaag atatgcagct ggtgcatttt gacttcaata 660
 gcacccatt cgagctgcac gacgactcat acgtcctaaa agaaatgcga ttctgtggaa 720
 gagaacaaca tgacagcact caagagtcga tatcaaatgg aggtgagagc tcaagcgtgt 780
 caaatattta ttggcaacac gtacagcatg tgaaaaggcc aaacaatggc acaggtaggt 840
 taccgagctc accgcctatt ccagggtatc tgaaaaggcg tgcgaagcac gagcactaac 900
 gctaggggtg ttggtcactt tcttttctgt ctgagcagtt ggttttattg catttctccg 960
 ttgtgtaaag ggccctgtaa attattaggg aagcgggagc gtagcttgat ctaatttagc 1020
 tctgcaccag attggtagaa cgacgggtgc tctaagtagt tgtgtaacta taactatcct 1080
 tgaatcgggt tcttttagtat ggttgagaga agggttgaca tgtaattttg tagagccttg 1140
 taaaattaaa attgttgatt tctatcaggg tgaaatttct gggcacaaaa aaa 1193

<210> 12
 <211> 194

<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (42)

<400> 12
Asp Lys Gly Asn Arg Gly Leu Arg Gln Phe Ser Met Lys Val Cys Glu
1 5 10 15
Lys Val Glu Ser Lys Gly Arg Thr Thr Tyr Asn Glu Val Ala Asp Glu
20 25 30
Leu Val Ala Glu Phe Thr Asp Pro Asn Xaa Asn Ile Glu Ala Asn Ala
35 40 45
Gln Gln Tyr Asp Glu Lys Asn Ile Arg Gly Arg Val Tyr Asp Ala Leu
50 55 60
Asn Val Leu Met Ala Met Asp Ile Ile Ser Lys Asp Lys Lys Glu Ile
65 70 75 80
Gln Trp Lys Gly Leu Pro Arg Thr Ser Ile Ser Asp Ile Glu Glu Leu
85 90 95
Lys Thr Glu Leu Val Gly Leu Lys Gly Arg Ile Glu Lys Lys Ser Val
100 105 110
Tyr Leu Gln Glu Leu Gln Asp Gln Tyr Val Gly Leu Gln Asn Leu Ile
115 120 125
Gln Arg Asn Gln Arg Asn Glu Gln Leu Tyr Gly Ser Gly Asn Thr Pro
130 135 140
Ser Gly Gly Val Ala Leu Pro Phe Ile Leu Val Gln Thr Arg Pro His
145 150 155 160
Ala Thr Val Glu Val Glu Ile Ser Glu Asp Met Gln Leu Val His Phe
165 170 175
Asp Phe Asn Ser Thr Phe Glu Leu His Asp Asp Ser Tyr Val Leu Lys
180 185 190
Glu Met

<210> 13
<211> 698
<212> DNA
<213> Triticum aestivum

<400> 13
gcacgaggtt tgagtgatat tgataaattg aagactgagg tcattgggct gaaaggtaga 60
attgacaaga aaagtgcata tctgcaggaa ttacaagatc aatatgcggg cctccaaaat 120
ttggtagagc gaaatgagca gctatatggt tcgggagatg ctccatctgg cggagtggcc 180
ctgccattca tattggttca gacacgtcct catgcaactg tcgaagtgga gatatcagaa 240
gatatgcagt tgggtgcattt tgatttcaat agcactccgt ttgagttgca cgatgattcc 300
tttgtattga aagcaatggg gttctctggt aaagaagaaa ctgacgggtac agtggctctg 360
gttgcaaatg cgggtgaaat ctcaagtgca tcaaatgttt atgggcgtcg atcaccacaa 420
cctgcaaggc caaatggaat taggctacga acctcacctc ctattccagg gatactgaaa 480
gggcgtgtca agcatgaaca ctaggcacta aaatgggttg ttaatgtttt gagagctact 540
tggatttata ctcccttggc ttctgtaact taacatgtaa aaaggcttgt aaattatgtc 600

catggggaaa tgtgatgttc taatttagct ttgcagccga tggtaggctg atcggcattgt 660
gaagctccag gcctgatttt aacttttagc tcaaaaaa 698

<210> 14
<211> 92
<212> PRT
<213> Triticum aestivum

<400> 14
Arg Ile Asp Lys Lys Ser Ala Tyr Leu Gln Glu Leu Gln Asp Gln Tyr
1 5 10 15
Ala Gly Leu Gln Asn Leu Val Glu Arg Asn Glu Arg Asn Glu Gln Leu
20 25 30
Tyr Gly Ser Gly Asp Ala Pro Ser Gly Gly Val Ala Leu Pro Phe Ile
35 40 45
Leu Val Gln Thr Arg Pro His Ala Thr Val Glu Val Glu Ile Ser Glu
50 55 60
Asp Met Gln Leu Val His Phe Asp Phe Asn Ser Thr Phe Glu Leu His
65 70 75 80
Asp Asp Ser Phe Val Leu Lys Ala Met Gly Phe Ser
85 90

<210> 15
<211> 540
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (413)

<220>
<221> unsure
<222> (461)

<220>
<221> unsure
<222> (471)

<220>
<221> unsure
<222> (498)

<220>
<221> unsure
<222> (517)

<400> 15
ctgaagatga catcaagtct ttgccctgct tccaagaatc agacactgat cgcaatcaaa 60
gcgcctcatg gtacaacttt ggagggtccca gacccctgatg aagtgaatga ttatccacag 120
aggagatata ggattgttct aaggagtact atgggtccaa tagatgtcta cctagttagt 180
caatttgagg agatgagtgg catggagact cctccaagga ctgtgcagcc agtaagcatg 240
gattctctag agaatcccag gacgccattg gctgcagaac ccaacaaagc tgcagagtca 300
caagccaaat attcaagatg ggctgctaatt ggcccttctga tgctccttct agttcacaag 360
gatattggcg ggatgattga agaattgtcc cttcagaaac tttgataccg atngcagact 420
actgggctcc ttatcaaaat tgccgggggg ttagcaatta nccgacatgt ngggaagaac 480
aagcaacaaa aaggtggnng tggggaaaagg gaatccnaaa aaattcaatt gcaagagggg 540

<210> 16
 <211> 53
 <212> PRT
 <213> Oryza sativa

<220>
 <221> UNSURE
 <222> (31)

<400> 16
 Leu Cys Pro Ala Ser Lys Asn Gln Thr Leu Ile Ala Ile Lys Ala Pro
 1 5 10 15
 His Gly Thr Thr Leu Glu Val Pro Asp Pro Asp Glu Val Asn Xaa Gln
 20 25 30
 Arg Arg Tyr Arg Ile Val Leu Arg Ser Thr Met Gly Pro Ile Asp Val
 35 40 45
 Tyr Leu Val Ser Gln
 50

<210> 17
 <211> 1600
 <212> DNA
 <213> Glycine max

<400> 17
 gtacgagcga atcttcttca ttacatgggt tcttcagatc ccatctcttc tcgccactac 60
 acttacaaca gaaaacaaaa atctctaggg cttctctgca ctaatttctt gagcttgtag 120
 aaccgagaca ccgttcactt gattggccta gacgatgctg ccaccgatt aggtgttgag 180
 agacggcgga tctatgacat tgtgaatgtc ctggagagta tcgggggtact ctcaaggaaa 240
 gcaaaaaatc aatacacatg gagagggttt gccgcaatcc ctctcactct acaggatctc 300
 aaggaagagg ggttgaagga gaatttctaatt tccctacgtg ggcttgaaa ccatgataaa 360
 gtttccgacg acgaagatga cgaggaaaca cagtccaatc ccgccgtac tgggaagtcaa 420
 agtgacaagt tgaaccctaa ttctactctt cctaaacctc tgaaaaatga aaatcgaaga 480
 gaaaaatctc tggctctgct taccagaat tttgtcaagc tctttgtctg ctctaattgtg 540
 gaattgatct cctttgacga ggctgcaaaa ttgctacttg gggatgcccc taatacgtct 600
 gtaatgagaa ctaaagttag acgcctatat gacatcgcaa atgtgctatc ctccatgaac 660
 cttattgaga agaccatac aatggatata cgaaaaccag cattcaggtg gctcggttct 720
 gaagggaaga catgggatga gacacttcac aaatcaaac taaacgactc taggaaaagg 780
 gcgtttggaa gtgatatac taacatcagt tttagagaga ataaagtggg attgttcacg 840
 agtggggact tgaatccgaa tcttaagaag ccaagaatgg aaaatggtag tgggctgggt 900
 gaggcagacg aaaacaattt aaaacaaggc ataaaacaag cttcaaagag ctatgaattt 960
 ggtcctttcg ctccagcttg cgtgccccaa gttggagcct ctcaagaata taatatgaag 1020
 caggtgcatg actggggcag tcttgcaact gcacatagcc ctcaagtatca aaatgaagcc 1080
 ttgagagaa tcttctccca ttacatggaa tatgccaaag tcaaaactata ggtatgtttg 1200
 aggaagaggc cattacaaat tttgtaaacg tttagagaga gcaagacttg ggatactagt 1260
 gatttagtaa aatttagtct ggaaacttg tttagagaga taactttcct cactagagaa 1320
 gtaaaacttaa gtttgcaaat tttaatccaa acatgggggt taactttcct cactagagaa 1380
 tatctgtagt atgcattctc ttcatgcgtt ttagtagctg agcttccatt tactgccttt 1440
 tcaaccgtga tgttgaccga ctgaatcagc ttaggcagtt cttgcaaatt atttatcaca 1500
 tttgtttgtt gtatagtttc taccctttat ttaatgtagc aggaaatctt gtgcactact 1560
 cctgtactgc agtctggagc tgctaattta aggaaaataa aattcgagtt gttttaaaaa 1600
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

<210> 18
 <211> 80
 <212> PRT
 <213> Glycine max

<400> 18
 Met Ala Ser Ser Asp Pro Ile Ser Ser Arg His Tyr Thr Tyr Asn Arg
 1 5 10 15

Lys Gln Lys Ser Leu Gly Leu Leu Cys Thr Asn Phe Leu Ser Leu Tyr
20 25 30
Asn Arg Asp Thr Val His Leu Ile Gly Leu Asp Asp Ala Ala Thr Arg
35 40 45
Leu Gly Val Glu Arg Arg Arg Ile Tyr Asp Ile Val Asn Val Leu Glu
50 55 60
Ser Ile Gly Val Leu Ser Arg Lys Ala Lys Asn Gln Tyr Thr Trp Arg
65 70 75 80



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/54, 9/12, 1/21, C07K 14/415, C12Q 1/68, 1/48		A3	(11) International Publication Number: WO 99/53075
			(43) International Publication Date: 21 October 1999 (21.10.99)
(21) International Application Number: PCT/US99/07638		(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 8 April 1999 (08.04.99)		Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 60/081,132 9 April 1998 (09.04.98) US		(88) Date of publication of the international search report: 23 March 2000 (23.03.00)	
(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): KLEIN, Theodore, M. [US/US]; 2229 Rosewood Drive, Wilmington, DE 19810 (US). MORAKINYO, Layo, O. [US/US]; 216 Barrett Run Place, Newark, DE 19702 (US). ODELL, Joan, T. [US/US]; P.O. Box 826, Unionville, PA 19375 (US). SAKAI, Hajime [DE/US]; 105 Banbury Drive, Wilmington, DE 19803 (US).			
(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).			
(54) Title: CELL CYCLE REGULATORY PROTEINS CDC-16, DP-1, DP-2 AND E2F FROM PLANTS			
(57) Abstract <p>This invention relates to an isolated nucleic acid fragment encoding a cell cycle regulatory protein. The invention also relates to the construction of a chimeric gene encoding all or a portion of the cell cycle regulatory protein, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the cell cycle regulatory protein in a transformed host cell.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

INTERNATIONAL SEARCH REPORT

Application No
PCT/EP 99/07638

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/54 C12N9/12 C12N1/21 C07K14/415 C12Q1/68 C12Q1/48		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C07K C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 01425 A (MEDICAL RES COUNCIL ;THANGUE NICHOLAS BARRIE (GB)) 18 January 1996 (1996-01-18) the whole document	6-10, 21-26
X	WO 97 43647 A (MEDICAL RES COUNCIL ;THANGUE NICHOLAS BARRIE (GB); LUNA SUSANA DE) 20 November 1997 (1997-11-20) the whole document	11-15, 21-26
X	WO 94 10307 A (MEDICAL RES COUNCIL ;THANGUE NICHOLAS BARRIE (GB)) 11 May 1994 (1994-05-11) the whole document	26
A	the whole document	6-10, 16-25
-/--		
<div style="display: flex; justify-content: space-between;"> <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. </div>		
* Special categories of cited documents : <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"Z" document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search <div style="text-align: center; font-weight: bold;">17 January 2000</div>		Date of mailing of the international search report <div style="text-align: center; font-weight: bold;">03.02.00</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer <div style="text-align: center; font-weight: bold;">Oderwald, H</div>

Form PCT/ISA210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

ational Application No

PCT/US 99/07638

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMHTG 'Online! EMBL, Heidelberg, Germany AC/ID AB007651, 31 October 1997 (1997-10-31) NAKAMURA Y ET AL.: "Structural analysis of Arabidopsis thaliana chromosome 5" XP002127845 abstract</p>	16-19, 22-25
X	<p>----- DATABASE R57U036 'Online! EMBL, Heidelberg, Germany AC/ID AB010693, 3 February 1998 (1998-02-03) NAKAMURA Y: "Structural analysis of Arabidopsis thaliana chromosome 5" XP002127846 abstract</p>	16-19, 22-25
X	<p>----- VEYLDER DE L ET AL: "THE ARABIDOPSIS CKS1 AT PROTEIN BINDS THE CYCLIN-DEPENDENT KINASES CDC2AAT AND CDC2BAT" FEBS LETTERS, vol. 412, no. 3, 4 August 1997 (1997-08-04), pages 446-452, XP002047992 ISSN: 0014-5793 the whole document</p>	26
A		1-25
A	<p>----- LAMB J R ET AL: "CDC16P, CDC23P AND CDC27P FORM A COMPLEX ESSENTIAL FOR MITOSIS" EMBO JOURNAL, vol. 13, no. 18, 15 September 1994 (1994-09-15), pages 4321-4328, XP002030313 ISSN: 0261-4189 cited in the application. the whole document</p>	1-25
A	<p>----- COLASANTI J ET AL.: "Isolation and characterization of cDNA clones encoding a functional p34cdc2 homologue from Zea mays" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 88, April 1991 (1991-04), pages 3377-3381, XP002113194 the whole document</p> <p style="text-align: center;">----- -/-</p>	1-25

INTERNATIONAL SEARCH REPORT

Application No
PCT/US 99/07638

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No.
A	<p>MIAO G-H ET AL.: "Two functional soybean genes encoding p34cdc2 protein kinases are regulated by different plant developmental pathways"</p> <p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 90, February 1993 (1993-02), pages 943-947, XP002113195 the whole document</p>	1-25
A	<p>HASHIMOTO J ET AL.: "Isolation and characterization of cDNA clones encoding cdc2 homologues from Oryza sativa: a functional homologue and cognate variants"</p> <p>MOLECULAR AND GENERAL GENETICS, vol. 233, no. 1-2, May 1992 (1992-05), pages 10-16, XP002113196 the whole document</p>	1-25
A	<p>WO 97 47745 A (PELAYO SANZ BURGOS ANDRES ;XIE QI (ES); CONSEJO SUPERIOR INVESTIGA) 18 December 1997 (1997-12-18) see the claims page 1 -page 6; examples 1,3</p>	1-25
A	<p>DUDITS ET AL: "Cyclin-dependent and calcium-dependent kinase families: response of cell division cycle to hormone and stress signals"</p> <p>PORTLAND PRESS RESEARCH MONOGRAPH, 1996, pages 21-45, XP002080506 ISSN: 0964-5845 the whole document</p>	1-25
P,X	<p>DATABASE EMHTG 'Online! EMBL Heidelberg, Germany AC/ID AB019226, 16 November 1998 (1998-11-16) NAKAMURA Y: "Structural analysis of Arabidopsis thaliana chromosome 5" XP002127847 abstract</p>	11-14, 22-25
P,X	<p>DATABASE R58U022 'Online! EMBL, Heidelberg, Germany AC AI487429, 17 March 1999 (1999-03-17) ALCALA J ET AL.: "Generation of ESTs from tomato carpel tissue" XP002127848 abstract</p>	16-19, 22-25

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 99/07638

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Although claim 21 is directed to a method of treatment of the human/animal body (as far as in vivo methods are concerned), the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: (1-5 complete); (21-25 partially)

Isolated nucleic acid encoding CDC16 as in SEQ ID NO: 1 and 3, chimeric genes comprising said nucleic acid, host cells comprising said gene, CDC16 polypeptide as in SEQ ID NO: 2 and 4, method of altering the level of expression of a cell cycle regulatory protein, method of obtaining said nucleic acid.

2. Claims: (6-10 complete); (21-25 partially)

same as in invention 1 but comprising DP-1 and SEQ ID NO: 5-10.

3. Claims: (11-15 complete); (21-25 partially)

same as in invention 1 but comprising DP-2 and SEQ ID NO: 11-14.

4. Claims: (16-20 complete); (21-25 partially)

same as in invention 1 but comprising E2F and SEQ ID NO: 15-18.

5. Claim : (26 complete)

Method for evaluating compounds inhibiting a cell cycle regulatory protein.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/07638

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9601425 A	18-01-1996	AT 186119 T AU 698473 B AU 2803695 A CA 2193091 A DE 69513025 D DE 769144 T EP 0769144 A FI 965289 A JP 10502252 T NO 965584 A NZ 288712 A US 5871901 A	15-11-1999 29-10-1998 25-01-1996 18-01-1996 02-12-1999 10-07-1997 23-04-1997 31-12-1996 03-03-1998 28-02-1997 24-09-1998 16-02-1999
WO 9743647 A	20-11-1997	AU 2784997 A EP 0914613 A US 5859199 A	05-12-1997 12-05-1999 12-01-1999
WO 9410307 A	11-05-1994	AT 181360 T AU 5343994 A CA 2148258 A DE 69325383 D DE 69325383 T EP 0669976 A EP 0905236 A JP 8503128 T NO 951641 A NZ 257181 A US 5863757 A	15-07-1999 24-05-1994 11-05-1994 22-07-1999 25-11-1999 06-09-1995 31-03-1999 09-04-1996 29-06-1995 27-07-1997 26-01-1999
WO 9747745 A	18-12-1997	WO 9747647 A AU 3257997 A CA 2257828 A CN 1227605 A EP 0914436 A	18-12-1997 07-01-1998 18-12-1997 01-09-1999 12-05-1999

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.